**深 圳 大 学 实 验 报 告**

**课程名称： 算法设计与分析**

**实验名称： 多种排序算法的算法实现及性能比较**

**学院： 计算机与软件学院 专业： 计算机科学与技术**

**报告人： 张健哲 学号： 2013150372 班级： 3**

**同组人： 无**

**指导教师： 李炎然**

**实验时间： 2015/3/25——2015/4/8**

**实验报告提交时间： 2015/4/8**

**教务处制**

**一．实验目的**

1. 掌握选择排序、冒泡排序、合并排序、快速排序、插入排序算法原理
2. 掌握不同排序算法时间效率的经验分析方法，验证理论分析与经验分析的一致性。

**二．实验步骤与结果**

**实验总体思路：**

利用switch结构来选择实验所要用的排序算法，每一种排序都用相同的计算运行时间的代码，不同的算法就在算法实现部分进行改动**（如下代码1至5所示）**。不断的改变数据规模，每一个规模在实验时，用循环进行多次实验并作为样本记录消耗的时间。最后输出在不同排序算法下，不同的数据规模的20次实验样本和平均用时**（如下图1至5所示）**。

**各排序算法的实现及实验结果：**

（注1：以下代码全部为伪代码，具体代码实现请参照程序中的代码）

（注2：图中显示的时间单位均为毫秒，图中“排序所花时间”一项为平均消耗时间，平均消耗时间结果以20组样本计算平均值后取整得到（并非四舍五入）。）

**1、选择排序**

代码1：

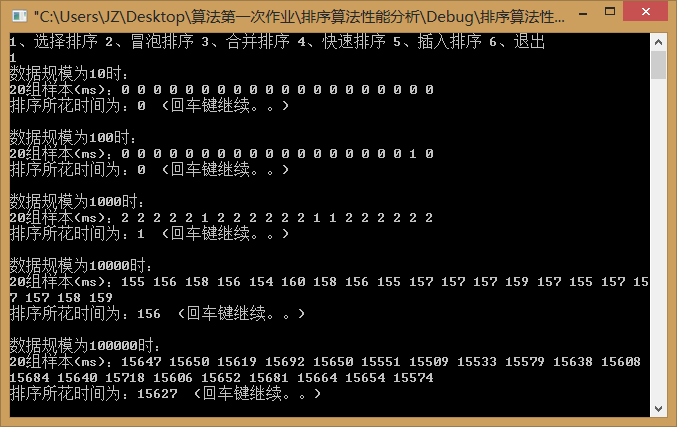
for i=0 to n-2

min=i

for j= i+1 to n-1

if ele[min]>ele[j] min=j

swap(ele[i],ele[min]) //交换



**图1、选择排序在不同数据规模下排序所消耗的时间**

**2、冒泡排序**

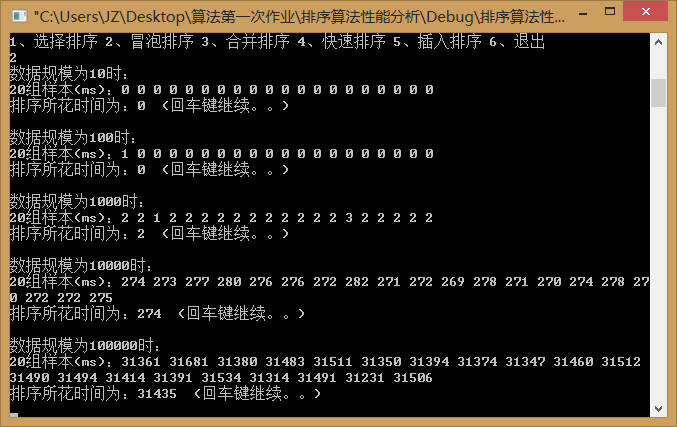
代码2：

for i= 0 to n-1

for j=0 to n-1-i

if a[j]>a[j+1]

swap(a[j],a[j+1]) //交换



**图2、冒泡排序在不同数据规模下排序所消耗的时间**

**3、合并排序**

代码3：

Merge(ele[1...n],left,right)

middle=(left+right)/2

if right>1eft+1

Merge(ele,left,middle)

Merge(ele,middle+1,right)

l←left r←right i←left

while l<=middle&&r<=right //两组分别一一比较，数据小的放入ele

if ele[l]<=ele[r]

t[i++]←ele[l++]

else

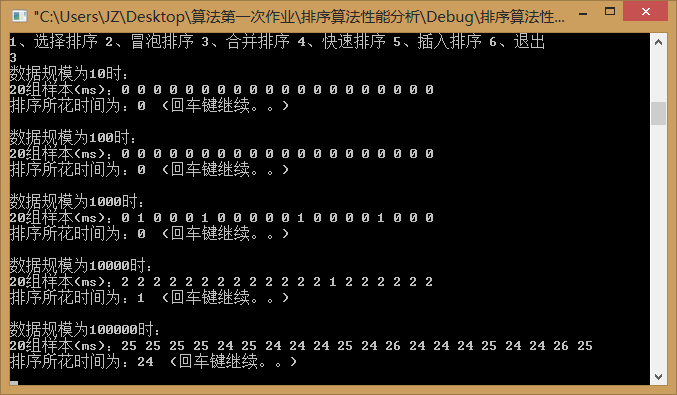
t[i++]←ele[r++]

while l>middle&&r<=r //只剩一组还有剩余的时，将剩下的按顺序放入

ele[i++]=s[r++]

while l<=middle && r>right

ele[i++]=s[l++];



**图3、合并排序在不同数据规模下排序所消耗的时间**

**4、快速排序**

代码4：

quick(ele[0...n-1],left,right)

if l<r

l←left r←right x←ele[l];

while l<r

while l<r && x<=ele[r] //找到一个比x小的数之后交换到前面的部分

r--

if l<r

ele[l]←ele[r] l++

while l<r && x>ele[l] //与上面相反

ll++

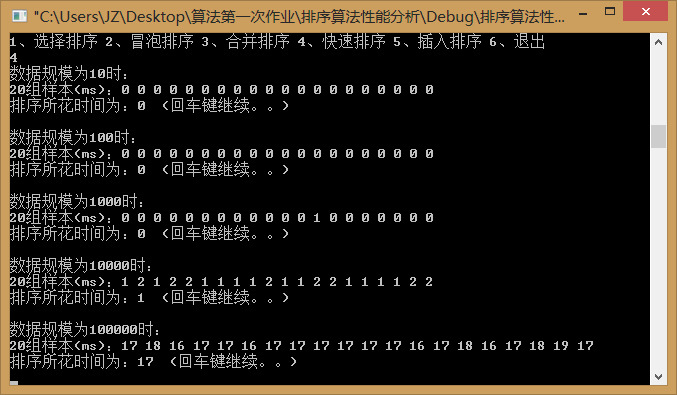
if l<r

ele[r]←ele[l] r--

ele[l]←x;

quick(ele,left,l-1) // 递归调用

quick(ele,l+1,right)



**图4、快速排序在不同数据规模下排序所消耗的时间**

**5、插入排序**

代码5：

for i=1→n-1

if ele[i]<ele[i-1] temp=ele[i]

for j= i-1 to 0 && ele[j]>temp

ele[j+1]←ele[j]

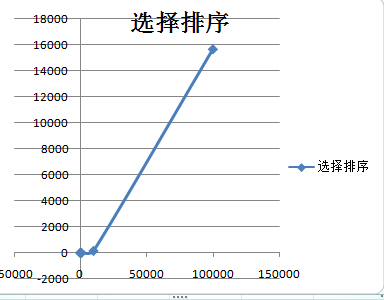
ele[j+1]←temp



**图5、插入排序在不同数据规模下排序所消耗的时间**

**三．实验分析**

**选择排序：**

****

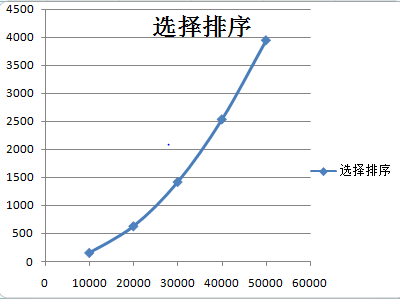
**图6、由图1数据整合而成的折线图**

为了更清晰的看到排序的数据规模与排序所需时间之间的影响，***我将实验的数据规模进行了一些调整，得到的平均数据依旧是以20组数据样本取平均数算得***（如下表1、图7所示）：

（由于图片占空间大且表达不直白，我将所得数据做成表格分析，下同）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 数据规模： | 10000 | 20000 | 30000 | 40000 | 50000 |
| 耗时（ms） | 158 | 634 | 1424 | 2541 | 3953 |

**表1、选择排序在不同数据规模下排序所消耗的时间**



**图7、由表1数据整合而成的折线图**

**图形上：**

形状基本符合n2（二次增长）

**数据上：**

我们发现当数据规模增大两倍时： 当数据规模增大两倍时：

10000→20000： 158\*22=632≈634 10000→30000：158\*32=1422≈1424

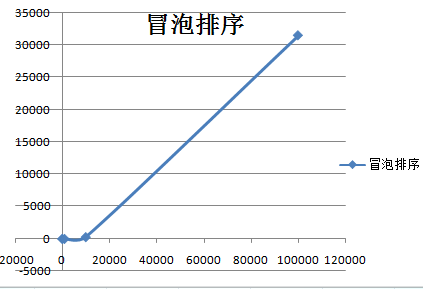
20000→40000： 634\*22=2536≈2541

其他倍数也可得到类似的结果。

**结论：**

我们发现，由于时间复杂度是o（n2）并且该排序的主要操作是比较操作，当数据规模扩大n倍时，相应的在时间的消耗上会扩大n2倍，同时我们发现，理论上乘以n2后的数据普遍会略小于实际数据，这主要原因可能是除了比较操作之外，赋值操作也随着n的增加逐渐增大，并且会在时间上体现出来，此外轻微的误差可能是数据的差异造成或者电脑等其他因素造成。

**冒泡排序：**

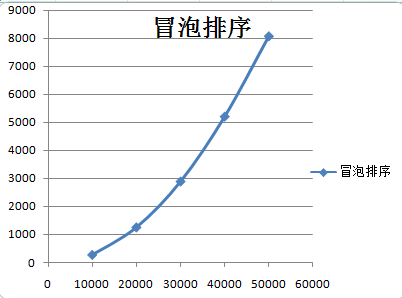
****

**图8、由图2数据整合而成的折线图**

为了更清晰的看到排序的数据规模与排序所需时间之间的影响，我将实验的数据规模进行了一些调整，得到的平均数据依旧是以20组数据样本取平均数算得（如下表2、图9所示）：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 数据规模： | 10000 | 20000 | 30000 | 40000 | 50000 |
| 耗时（ms） | 284 | 1266 | 2899 | 5213 | 8075 |

**表2、冒泡排序在不同数据规模下排序所消耗的时间**



**图9、由表2数据整合而成的折线图**

**图形上：**

形状基本符合n2（二次增长）

**数据上：**

我们发现当数据规模增大两倍时： 当数据规模增大两倍时：

10000→20000：284\*22=1136≠1266(误差130） 10000→30000：284\*32=2556≠2899(误差343）

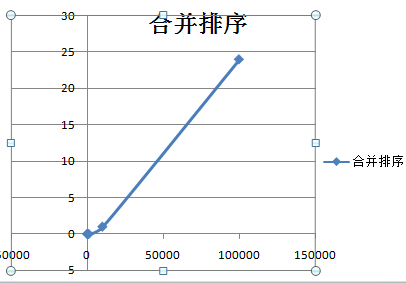
20000→40000:1266\*22=5064≠5313(误差149）

其他倍数也可得到类似的结果。

**结论：**

我们发现，虽然时间复杂度是o（n2），但当数据规模扩大n倍时，并没有相应的在时间的消耗上扩大n2倍，而是多于n2，同时我们发现，这个误差会随着数据规模的扩大而扩大，这主要原因是除了比较操作之外，赋值操作也随着n的增加逐渐增大，而且事实证明在数据比较极端的情况下，赋值操作已经不能忽略不计（这里的赋值操作发生在数据交换时所需要的操作），最糟糕的情况是每次比较就要进行三次的赋值操作，与此相比，电脑等其他因素造成轻微的误差可以忽略不计。

**合并排序：**

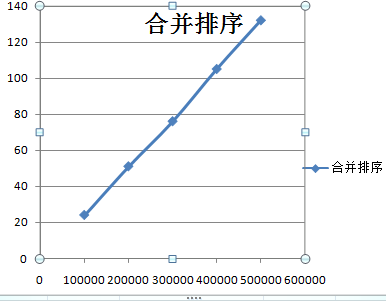
****

**图10、由图3数据整合而成的折线图**

为了更清晰的看到排序的数据规模与排序所需时间之间的影响，我将实验的数据规模进行了一些调整，得到的平均数据依旧是以20组数据样本取平均数算得（如下表3、图11所示）：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 数据规模： | 100000 | 200000 | 300000 | 400000 | 500000 |
| 耗时（ms） | 24 | 51 | 76 | 105 | 132 |

**表3、合并排序在不同数据规模下排序所消耗的时间**



**图11、由表3数据整合而成的折线图**

**图形上：**

形状基本符合n（线性增长）

**数据上：**

我们发现当数据规模增大两倍时： 当数据规模增大两倍时：

10000→20000： 24\*2log22=48≈51 10000→30000：24\*3log23=72≈76

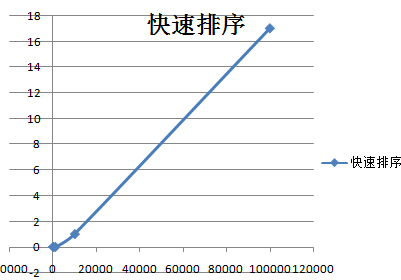
20000→40000： 51\*2log222=102≈105

其他倍数也可得到类似的结果。

**结论：**

我们发现，由于时间复杂度是o（nlog2n），当数据规模扩大n倍时，相应的在时间的消耗上会扩大nlog2n倍，同时我们发现，理论上乘以nlog2n后的数据普遍会略小于实际数据，这主要原因快速排序需要递归调用，递归调用需要花费额外的时间，此外轻微的误差可能是数据的差异造成或者电脑等其他因素造成。

**快速排序：**

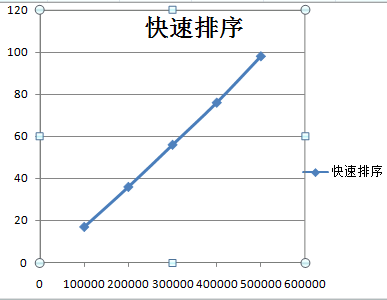
****

**图12、由图4数据整合而成的折线图**

为了更清晰的看到排序的数据规模与排序所需时间之间的影响，我将实验的数据规模进行了一些调整，得到的平均数据依旧是以20组数据样本取平均数算得（如下表4、图13所示）：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 数据规模： | 100000 | 200000 | 300000 | 400000 | 500000 |
| 耗时（ms） | 17 | 36 | 56 | 76 | 98 |

**表4、快速排序在不同数据规模下排序所消耗的时间**



**图13、由表4数据整合而成的折线图**

**图形上：**

形状基本符合n（线性增长）

**数据上：**

我们发现当数据规模增大两倍时： 当数据规模增大两倍时：

10000→20000： 17\*2log22=34≈36 10000→30000：17\*3log23≈56

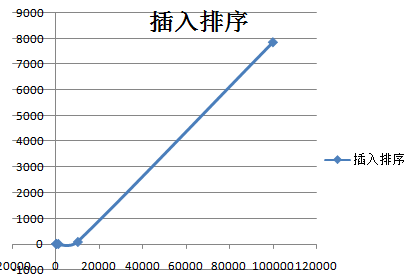
20000→40000： 26\*2log22=54≈56

其他倍数也可得到类似的结果。

**结论：**

我们发现，由于时间复杂度是o（nlog2n），当数据规模扩大n倍时，相应的在时间的消耗上会扩大nlog2n倍，同时我们发现，理论上乘以nlog2n后的数据普遍会略小于实际数据，这主要原因快速排序需要递归调用，递归调用需要花费额外的时间，此外轻微的误差可能是数据的差异造成或者电脑等其他因素造成。

**插入排序：**

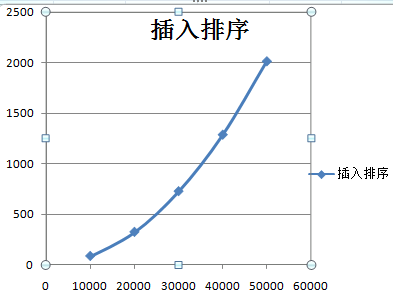
****

**图14、由图5数据整合而成的折线图**

为了更清晰的看到排序的数据规模与排序所需时间之间的影响，我将实验的数据规模进行了一些调整，得到的平均数据依旧是以20组数据样本取平均数算得（如下表5、图15所示）：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 数据规模： | 10000 | 20000 | 30000 | 40000 | 50000 |
| 耗时（ms） | 80 | 319 | 723 | 1283 | 2011 |

**表5、插入排序在不同数据规模下排序所消耗的时间**



**图15、由表5数据整合而成的折线图**

**图形上：**

形状基本符合n2（二次增长）

**数据上：**

我们发现当数据规模增大两倍时： 当数据规模增大两倍时：

10000→20000： 80\*22=320≈319 10000→30000：80\*32=720≈723

20000→40000： 319\*22=1276≈1283

其他倍数也可得到类似的结果。

**结论：**

我们发现，由于时间复杂度是o（n2），当数据规模扩大n倍时，相应的在时间的消耗上会扩大n2倍，理论上，如果数据太具有特殊性，那此算法被影响的程度会比较大，他的的比较次数可以从线性次数，到n2次，赋值次数也可能由常数次变成n2总的来说，受数据影响较大，由于我实验基数少，所以无法得出此结论。

。

。

。

。

。

。

。

。

。

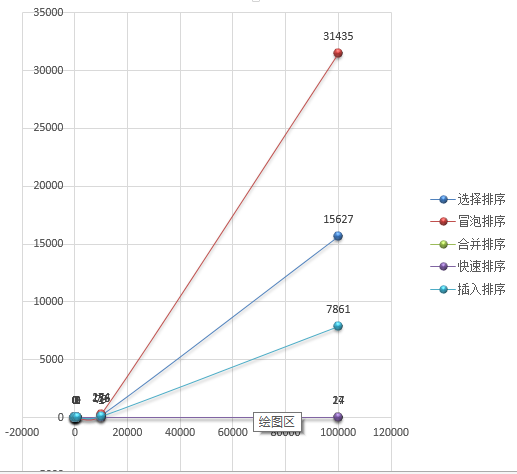
。

。

。

。

将五种排序的实验汇总在一起，如下图16所示



**图16、由图6、8、10、12、14整合而来**

从图中以及之前的分析中我们可以得到以下结论

时间复杂度同样为o（n2）的选择、冒泡和合并排序，在数据处理上相差的也比较大，其中1、冒泡排序平均耗时最多，**其主要原因是：**冒泡排序在比较次数上达到了o（n2），但这种排序同时也受交换次数的影响，而且最多时间复杂度也是o（n2）。如此，同样是o（n2），但冒泡排序的二次项系数会比另外两个大不少，所以最为耗时。 2、选择排序和插入排序相比较，插入排序更快，**其原因是：**选择排序需要从序列中找到当前最大或最小的值才能进行排序，因此每次都需要与子序列中的全部元素进行比较。而插入排序无需比较子序列全部元素，只需要找到当前序列第一个比自己大或小的元素，将自身插入到其前一个位置即可。3、同样是o（nlog2n）但快速排序更快：快速排序出现最差的情况并不是由于输入数据，而是选取到的随机数本身，选到极端的情况非常小，所以对于绝大部分数据而言都是能达o（nlog2n），而合并排序需要赋值的语句还是较多，受输入数据的影响比快排大，所以当数据规模较大时，不受输入数据影响的快速排序更快

**四．实验心得**

本次实验虽然花费很大的心思，但确实让我对这几种排序的认识更加深刻，同样的数据，排序的时间可以相差如此之大，这可能会改变我每次都使用冒泡排序的这一习惯，同时，对算法的优良性不同而导致的结果差异之大，感觉到好的算法是多么的重要，当然合理利用算法也是不可忽视的。这次实验虽然花了很大精力，却收获累累。

|  |
| --- |
| 指导教师批阅意见：  成绩评定：  指导教师签字：  年 月 日 |
| 备注： |

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。

2、教师批改学生实验报告时间应在学生提交实验报告时间后10日内。